Network Working Group
Request for Comments: 2796
Updates: 1966
Category: Standards Track

T. Bates
Cisco Systems
R. Chandra
E. Chen
Redback Networks

April 2000

BGP Route Reflection – альтернатива полносвязности IBGP

BGP Route Reflection - An Alternative to Full Mesh IBGP

Статус документа

В этом документе содержится спецификация протокола, предложенного сообществу Internet. Документ служит приглашением к дискуссии в целях развития и совершенствования протокола. Текущее состояние стандартизации протокола вы можете узнать из документа «Internet Official Protocol Standards» (STD 1). Документ может распространяться без ограничений.

Авторские права

Тезисы

BGP[1] [1] представляет собой протокол междоменной маршрутизации, разработанный для сетей TCP/IP. В настоящее время в сети Internet протокол BGP настроен так, что все узлы BGP в одной AS должны образовывать полносвязный набор соединений (fully meshed) и любая внешняя маршрутная информация должна передаваться всем остальным маршрутизаторам внутри данной AS. Это порождает серьезные проблемы масштабирования, которые подробно описаны вместе с альтернативными предложениями в документах [2,3].

В данном документе описан метод «отражения маршрутов» (Route Reflection) и его использование, ослабляющее требование полносвязности для IBGP.

1. Введение

В настоящее время в сети Internet протокол BGP настроен так, что все узлы BGP в одной AS должны образовывать полносвязный набор соединений и любая внешняя маршрутная информация должна передаваться всем остальным маршрутизаторам внутри данной AS. Для n узлов BGP в данной AS требуется организовать n*(n-1)/2 уникальных сессий IBGP. Очевидно, что требование полносвязности становится невыполнимым в системах, где большое число узлов IBGP обменивается значительными объемами маршрутной информации (такая ситуация наблюдается в большинстве современных сетей).

Эта проблема масштабирования и многочисленные предложения по снижению ее остроты подробно описаны в документах [2,3]. Данный документ представляет еще один вариант избавления от полносвязности, известный как Route Reflection. Этот метод позволяет узлу BGP (называемому Route Reflector) анонсировать полученные от IBGP маршруты некоторым партнерам IBGP. Он изменяет общепринятую концепцию работы и добавляет два новых необязательных непереходных[2] атрибута BGP для предотвращения петель при обновлении маршрутов.

Данный документ является пересмотром RFC1966 [4] и включает редакционные правки, пояснения и корректировки, основанные на опыте использования отражения маршрутов. Список изменений приведен в Приложении.

2. Базовые требования

Метод Route Reflection удовлетворяет перечисленным ниже критериям.

- *Простота*

  Любое дополнение должно быть понятным и простым в настройке.

- *Простота перехода*

  Должна обеспечиваться возможность перехода от полносвязной конфигурации без необходимости изменения топологии или AS. Метод, предложенный в [3], вносит слишком высокие издержки в части управления.

- *Совместимость*

- Должна обеспечиваться возможность сохранения не поддерживающих данный метод узлов IBGP как части исходной AS или домена без потери какой-либо маршрутной информации BGP.

---

[1] Border Gateway Protocol – протокол граничного шлюза.
[2] В оригинале ошибочно сказано про два переходных атрибута, что не соответствует определениями главы 7. *Прим. перев.*

Эти критерии основаны на опыте использования метода в очень больших сетях со сложной топологией и множеством внешних соединений.

3. Отражение маршрутов

Основная идея метода отражения (Route Reflection) очень проста. Рассмотрим пример, показанный на рисунке 1.

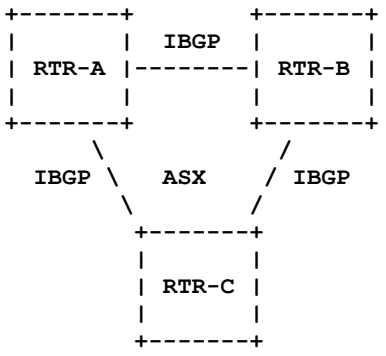


Рисунок 1: Полносвязная система IBGP

В автономной системе ASX имеется три узла IBGP (маршрутизаторы RTR-A, RTR-B, RTR-C). В рамках существующей модели BGP если RTR-A получает внешний маршрут и выбирает этот маршрут в качестве лучшего, он должен анонсировать этот внешний маршрут обоим узлам RTR-B и RTR-C. Узлы RTR-B и RTR-C (как узлы IBGP) не будут заново анонсировать этот полученный от IBGP маршрут другим партнерам IBGP.

Если это правило ослабить и позволить узлу RTR-C анонсировать полученные от IBGP маршруты другим партнерам IBGP, тогда он будет реанонсировать (или отражать) маршруты IBGP, полученные от RTR-A, узлу RTR-B и наоборот. Это позволит отказаться от организации сессии IBGP между узлами RTR-A и RTR-B, как показано на рисунке 2.

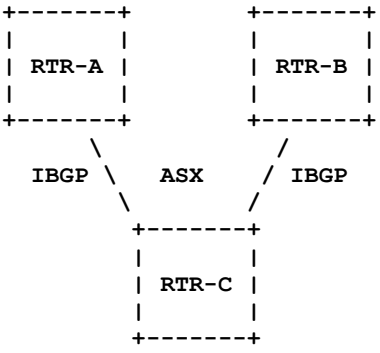


Рисунок 2: IBGP с отражением маршрутов

Схема метода Route Reflection основана именно на этом принципе.

4. Терминология и концепции

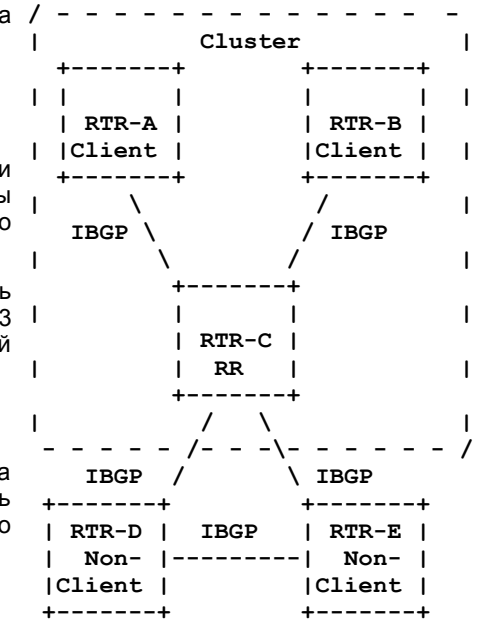
Мы используем термин «отражение маршрутов» для описания действий узла BGP, анонсирующего полученные от IBGP маршруты другим партнерам IBGP. Если об узле BGP говорят как об «отражателе маршрутов» (Route Reflector или RR), это означает, что данный узел «отражает» полученные маршруты своим партнерам.

Внутренние партнеры узла RR делятся на две группы:

1) Партнеры-клиенты.

2) Партнеры, не являющиеся клиентами.

Узел RR отражает маршруты между этими группами и может отражать маршруты между клиентами. Узел RR вместе со своими клиентами образует кластер (Cluster).



Партнеры, не являющиеся клиентами (Non-Client peer), должны сохранять полносвязность, но для клиентов это требование снимается. На рисунке 3 показан пример сети с базовыми компонентами RR, иллюстрирующий терминологию.

Рисунок 3: Компоненты RR

5. Работа метода

Когда RR получает маршрут от партнера IBGP, он выбирает лучший путь на основе своих критериев. После выбора лучшего пути узел должен выполнить перечисленные ниже операции в зависимости от типа партнера, передавшего информацию о лучшем пути:

1) Маршрут получен от партнера, не являющегося клиентом.

   Отразить маршрут всем клиентам.

2) Маршрут получен от клиента.

   Отразить маршрут всем партнерам, не являющимся клиентами, а также партнерам-клиентам (поскольку клиенты могут не быть полносвязными).

Автономная система может включать множество RR. Узел RR трактует остальные рефлекторы RR, как обычные внутренние узлы BGP. Рефлектор RR может быть настроен на присутствие других RR как в числе клиентов, так и среди партнеров, не являющихся клиентами.

В простой конфигурации опорная сеть может быть поделена на множество кластеров. Каждый рефлектор RR настраивается на то, что другие RR не относятся к группе клиентов (таким образом, все RR будут образовывать полносвязную систему). Клиенты будут настраиваться на поддержку сессий IBGP только с RR в своем кластере. Благодаря отражению маршрутов все узлы IBGP будут получать отраженную маршрутную информацию.

В автономной системе могут присутствовать узлы BGP, не понимающие концепцию отражения маршрутов (будем называть их обычными узлами BGP). Схема отражения маршрутов допускает сосуществование с обычными узлами BGP. Такие узлы могут относиться к группе клиентов или не являться клиентами RR. Это обеспечивает возможность простого и постепенного перехода от существующей модели работы IBGP к модели с отражением маршрутов. Можно начать с создания кластера путем настройки одного маршрутизатора в качестве означенного RR и настройки остальных RR и их клиентов как нормальных партнеров IBGP. Постепенно могут создаваться дополнительные кластеры.

6. Избыточные RR

Обычно кластер клиентов будет включать один рефлектор RR. В этом случае кластер будет идентифицироваться значением ROUTER_ID рефлектора RR. Однако такой вариант может не обеспечивать достаточной надежности и для резервирования в одном кластере может создаваться множество RR. Все рефлекторы RR одного кластера могут быть

настроены на использование общего 4-байтового идентификатора CLUSTER_ID, который позволяет любому рефлектору RR отбрасывать маршруты, получаемые от других RR того же кластера.

7. Предотвращение петель

При использовании отражения маршрутов возможно возникновение петель при реанонсировании в результате некорректной конфигурации. Метод Route Reflection определяет два новых атрибута для детектирования и предотвращения таких петель.

ORIGINATOR_ID

ORIGINATOR_ID – необязательный, непереходный атрибут BGP с кодом типа 9. Этот атрибут имеет размер 4 байта и создается рефлектором RR в отраженном маршруте. Атрибут будет включать значение ROUTER_ID источника маршрута (originator) в локальной AS. Узлу BGP не следует создавать атрибут ORIGINATOR_ID, если последний уже присутствует. Маршрутизатору, распознающему атрибут ORIGINATOR_ID, следует игнорировать маршрут, содержащие значение его ROUTER_ID в качестве ORIGINATOR_ID.

CLUSTER_LIST

CLUSTER_LIST – необязательный, непереходный атрибут BGP с кодом типа 10. Этот атрибут представляет собой последовательность значений CLUSTER_ID, представляющих путь отражения, по которому передавался маршрут. Формат атрибута показан ниже.

```
         0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Attr. Flags  |Attr. Type Code|   Length      | value ...
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Поле Length[1] указывает число октетов.

Когда рефлектор RR отражает маршрут, он должен добавить локальное значение CLUSTER_ID в начало (prepend) CLUSTER_LIST. Если список CLUSTER_LIST пуст, узел должен создать новый список. Используя этот атрибут, RR может определять возникновение петель при передаче маршрутной информации в результате конфигурационных ошибок. Если локальное значение CLUSTER_ID присутствует в списке кластеров, полученный анонс следует игнорировать.

8. Вопросы реализации метода

Следует принять меры по предотвращению изменения описанных выше атрибутов пути (средствами конфигурации) в процессе обмена маршрутной информацией между RR и клиентами или партнерами, не являющимися клиентами. Такое изменение атрибутов может приводить к возникновению маршрутных петель.

Кроме того, когда RR отражает маршрут, ему не следует изменять в маршруте значения атрибутов NEXT_HOP, AS_PATH, LOCAL_PREF и MED, поскольку это может приводить к возникновению маршрутных петель.

9. Вопросы настройки и развертывания

Протокол BGP не обеспечивает клиентам способа динамической идентификации себя в качестве клиентов RR. Простейшим способом такой идентификации является настройка конфигурации вручную.

Одним из ключевых моментов метода отражения маршрутов в контексте проблемы масштабирования является то, что RR обрабатывает полученную информацию и отражает только лучший путь.

На выбор маршрута BGP могут оказывать влияние обе метрики MED и IGP. Поскольку атрибуты MED не всегда совместимы, а метрика IGP может отличаться для каждого маршрутизатора, в некоторых вариантах топологии отражения метод отражения может давать при выборе маршрута результат, отличающийся от случая полносвязной системы IBGP. Для получения совпадающих результатов в случаях использования отражения и полносвязной системы IBGP следует сделать так, чтобы рефлекторы маршрутов никогда не выбирали лучший маршрут BGP на основе метрики IGP, которая существенно отличается от IGP-метрики их клиентов, или на основе несовместимых атрибутов MED. Первый вариант может быть достигнут путем настройки конфигурации таким образом, чтобы внутрикластерная метрика IGP всегда давала преимущество перед межкластерной метрикой IGP, и поддержки полной связности (full mesh) внутри кластера. Для реализации второго варианта можно использовать любой из перечисленных ниже способов:

- устанавливать на граничном маршрутизаторе уровень локального предпочтения маршрутов в соответствии с MED;
- обеспечить, чтобы длины AS_PATH для разных AS различались при использовании длины пути в качестве критерия выбора;
- настроить основанную на группах (community) политику, использование которой позволит рефлектору выбрать лучший путь.

Можно утверждать, что второй вариант вносит чрезмерные ограничения и в некоторых случаях будет непрактичным. Можно также утверждать, что при отсутствии маршрутных петель не существует жесткой необходимости обеспечивать совпадение результатов выбора маршрута с использованием отражения и полносвязной системы IBGP.

Для предотвращения маршрутных петель и поддержки согласованной картины маршрутизации важно аккуратно рассмотреть топологию сети при выборе топологии отражения маршрутов. В общем случае топологию отражения следует делать конгруэнтной топологии сети, когда существует множество путей для данного префикса. Общепринятым является использование отражения на базе POP, при котором каждая точка POP поддерживает свои рефлекторы маршрутов, обслуживающие клиентов POP, и все рефлекторы образуют между собой полносвязную

[1]Поле Length ошибочно указано на рисунке, как однооктетное. Размер этого поля может составлять 1 или 2 октета в зависимости от значения флага Extended Length (см. параграф 4.3 RFC 4271). *Прим. перев.*

систему. В дополнение к этому клиенты рефлекторов в каждой POP зачастую также образуют полносвязную систему в целях оптимальной маршрутизации внутри POP, а внутренняя (для POP) метрика IGP является предпочтительной по сравнению с метрикой inter-POP IGP.

10. Вопросы безопасности

Это расширение протокола BGP не изменяет состояния безопасности, присущего IBGP [5].

11. Благодарности

Авторы благодарят Dennis Ferguson, John Scudder, Paul Traina и Tony Li за дискуссии, которые привели к созданию этого документа. Идея метода основана на давней дискуссии между Tony Li и Dimitri Haskin.

Кроме того, авторы хотят поблагодарить Yakov Rekhter за просмотр документа и полезные предложения, а также отметить полезные комментарии Tony Li, Rohit Dube и John Scudder к главе 9 и комментарии Bruce Cole.

13. Литература

[1] Rekhter, Y. and T. Li, "A Border Gateway Protocol 4 (BGP-4)", RFC 1771[1], March 1995.

[2] Haskin, D., "A BGP/IDRP Route Server alternative to a full mesh routing", RFC 1863, October 1995.

[3] Traina, P., "Autonomous System Confederations for BGP"[2], RFC 1965[3], June 1996.

[4] Bates, T. and R. Chandra, "BGP Route Reflection An alternative to full mesh IBGP", RFC 1966[3], June 1996.

[5] Heffernan, A., "Protection of BGP Sessions via the TCP MD5 Signature Option", RFC 2385[3], August 1998.

14. Адреса авторов

Tony Bates

Cisco Systems, Inc.

170 West Tasman Drive

San Jose, CA 95134

EMail: tbates@cisco.com

Ravi Chandra

Redback Networks Inc.

350 Holger Way.

San Jose, CA 95134

EMail: rchandra@redback.com

Enke Chen

Redback Networks Inc.

350 Holger Way.

San Jose, CA 95134

EMail: enke@redback.com

Перевод на русский язык

Николай Малых

nmalykh@protocols.ru

Приложение. Сравнение с RFC 1966

Разъяснены некоторые термины, связанные с отражением маршрутов и исключены упоминания маршрутов и узлов EBGP.

Разъяснена и сделана более согласованной обработка получателем маршрутных петель (в результате отражения).

Способ добавления атрибута CLUSTER_ID в список CLUSTER_LIST был заменен с «append» на «prepend» в соответствии с реализованным кодом.

В главу «Вопросы настройки и развертывания» добавлено рассмотрение некоторых вопросов развертывания метода.

Полное заявление авторских прав

[1]Этот документ утратил силу и заменен RFC 4271. Перевод имеется на сайте www.protocols.ru. *Прим. перев.*
[2]В оригинале этот документ ошибочно назван «Limited Autonomous System Confederations for BGP». *Прим. перев.*
[3]Перевод этого документа имеется на сайте www.protocols.ru. *Прим. перев.*

Подтверждение

Финансирование функций RFC Editor обеспечивается Internet Society.